

Book review

Diachronic Corpus Pragmatics

Irma Taavitsainen, Andreas H. Jucker and Jukka Tuominen (Eds.), John Benjamins, Amsterdam, 2014, 335 pp., ISBN: 9789027256485 (hardbound), EUR 95.00



Research into pragmatic phenomena from a diachronic perspective has always relied on two seemingly distant approaches. On the one hand, a pragmaticist could engage in qualitative data assessments, requiring a more philological treatment of the sources, usually distant historically and embedded in a specific socio-cultural milieu; on the other hand, more quantitative, corpus-based studies have also been pursued, to meet the requirements of replicability and to add statistical precision. The realization that it is the combination of these two approaches which provide the most informative and reliable insights into various pragmatic phenomena could not be expressed in a more pronounced manner than in the volume under review. The editorial trio, Irma Taavitsainen, Andreas H. Jucker and Jukka Tuominen, have brought forward a timely collection of studies which show the modern face of diachronic pragmatic studies.

The book originates from a thematic panel at the 12th International Pragmatics Conference in Manchester in 2011 and comprises chapters based on diachronic pragmatic developments in English, Dutch, Swedish, Italian, Spanish, Finnish, Estonian and Japanese. Bringing languages other than English into the picture is an undeniable asset of this enterprise. Far from a random selection, the chapters form a very coherent structure: from inquiries based on individual words, through larger units, to dialogues, across a wide linguistic typological background.

The book is a truly interdisciplinary endeavour. In their introduction, Andreas H. Jucker and Irma Taavitsainen show how historical linguistics, pragmatics and corpus studies, the three disciplines which at their inception were quite distant from each other, have increasingly combined into an exciting newly emerging discipline. The authors are careful to address several methodological challenges which characterize this highly complex approach to language change, e.g. editorial conventions, spelling variation or corpus annotation. The field of diachronic corpus pragmatics is driven by two objectives: expanding the database and getting more context. This is reflected in the employment of megacorpora, such as COHA, on the one hand, and small philological, often single-genre corpora, such as EMENT, on the other. Context is a central concept in a pragmatic inquiry. All contributions to the book aim to combine a high degree of contextualization with a quantitative assessment of emerging patterns. The attempt at balancing the two approaches is only very recent, which makes it a challenge also for the authors in this volume.

The first part of the book is devoted to “words”. The opening chapter in this section, by Claudia Claridge and Merja Kytö, is based on a subsection of the Old Bailey Corpus (1730s–1830s), which is vast enough (c. 49.5 million words) to allow a study of low-frequency phenomena in a diachronic perspective. The analysis addresses the differences in the scope and pace of grammaticalization between degree modifiers *a bit* and *pretty*. The authors give an overview of the semantic (the downscaling and upscaling), collocational and syntactic behaviour of these units and concentrate on the most prolific time for the general grammaticalizing trends in modifiers. One aspect is particularly worthy of attention: the phrase *a bit* shows a different palette of meanings, and, in fact, an increase in its function as a degree modifier when extended to *a bit of* and *a bit of a*. Lexical bundle research (e.g. Biber et al., 1999) has shown that repetition and formulaicity in language do not work within the rigid confines of linguistic units and full phrases. Claridge and Kytö’s paper shows that one has to separately consider the focal and the extended versions of what seems to be the same construction in a diachronic pragmatic study.

Another paper on grammaticalization of stance markers, by Taru Norlund and Heli Pekkarinen, concentrates on the Finnish adverbial *muka* ‘as if, supposedly, allegedly’. A very informative qualitative discussion of highly nuanced reportative and dubitative functions of *muka* is followed by an argument that this adverbial entered standard Finnish from eastern dialects and through specific genres. The authors juxtapose “data from dialects” and “written Finnish” in their study and point out that pragmatic research in Finland has relied on “an extensive use of dialect material alongside written texts” (p. 54). This distinction came as a surprise to me, because modern English-based linguistics does not treat the history of the standard as something disconnected from the history of other dialects, or as a more “legitimate” version of “the

language". This is a case of different linguistic traditions coming into contact, which is not a bad thing in itself, but it highlights the fact that caution is required in assuming the same theoretical backdrop in a volume based on contributions emerging from different scholarly traditions. On the basis of scientific texts (high in evidential contexts) and fairy tales (high in causal and explanatory contexts) Norlund and Pekkarinen show that the pragmaticalized use of *muka* is genre-dependent. In the chapter one also finds a highly interesting discussion on linguistic purism and nationalistic language planning in nineteenth-century Finland.

Rumiko Shinzato offers an intriguing analysis of pragmaticalization driven by the phonological similarity of several degree and manner adverbs in Japanese (enriched with some additional data from Korean). The data selection aims to reflect spoken discourse and is therefore rather patchy: one hundred years of dialogue passages from novels (1887–1986) and ten years of spontaneous conversation transcripts (1993–2003). Still, the author makes a persuasive case for the fluid boundary between speech and writing using Koch and Oesterreicher's framework of communicative immediacy and distance (pp. 81–82) and is able to trace a gradual loss of semantic transparency by the adverbs *amari*, *bakari* and *yahari* due to similar phonetic reductions. The proposed notion of "paradigmatic iconicity" is an elegant explanation for what is going on – different adverbs end up as pragmatic particles in a rather parallel fashion. Two more intriguing aspects of this chapter come forth: the impossibility of word count normalization in Japanese and the ordering of syntactic elements driven by pragmatics, with the propositional core expanding symmetrically left and right into subjective and intersubjective layers. Thus the position in a Japanese sentence is a rather reliable indicator of the pragmatic status of a given linguistic form.

Annika Küngas's chapter concentrates on the Estonian adverb *täpselt* 'exactly, precisely'. In the qualitative discussion the multiple meanings of *täpselt* are carefully illustrated, from the most concrete adverbial ones to those more elusive and purely discursive (e.g. emphatic). It seems that this polysemous adverb has undergone grammaticalization to a pragmatic marker (or pragmaticalization, [Claridge and Arnovick, 2010](#)), much along the same lines as described in literature for similar English grammaticalization scenarios, which points to a universal nature of the process. Still, looking at the quantitative part of the chapter, a diachronic change is not really borne out by the statistics. The ratio of adverbial and pragmatic uses of *täpselt* stays relatively constant throughout the dataset (1930s–2000s) and the diachronic dimension is missing from other correlations discussed (e.g. collocational patterns or the position in the clause). One should also wonder if it is justifiable to conflate all the adverbial meanings and contrast them sharply with pragmatic meanings (especially in the logistic regression tests), while the previous section emphasized potential ambiguity.

Two truth adverbs, *davvero* and *veramente*, are traced through "the whole documented history" of Italian (p. 133) by Davide Ricca and Jaqueline Visconti. The best feature of this chapter is that the authors do not shun difficult methodological questions, approach their material and methods with a critical eye, speculate and ask uncomfortable questions of their data. For instance, *davvero* experiences a rapid growth in the eighteenth century but the authors admit that this finding reflects the style of a particular playwright. The authors postulate that the syntactic position, especially in view of turn-taking patterns, may have played a role in different semantic and pragmatic developments of the two items. Furthermore, intersubjective meanings of mitigated rebuttal for *veramente* can be successfully studied only in a dialogic context, where the intonation also matters. The chapter proves that interactional factors may play a very important role in grammaticalization.

The second part of the book, devoted to larger syntactic units with a pragmatic function, opens with a descriptive study of multi-adjectival premodification in Early Modern English medical writing, by Jukka Tyrkkö. This is a very strong voice in favour of combining qualitative readings and well-grounded quantitative analyses. Stacked adjectives have been identified as a feature of present-day scientific texts, so the chapter traces this pattern across the sixteenth and seventeenth centuries – a time when the modern scientific style was being created. One of the most striking findings is a difference in the habits of non-professional medical writers, who used more adjectives, usually for emphasis and emotive appeal, and professional writers, who "showed more restraint" (p. 159). It turns out that adjectival patterns were sensitive to pragmatic motivations: marketing strategies, emotional appeal, and need for description. The study also shows an increase in complex adjectival premodification; interestingly, shorter adjectival combinations tend towards formulaic usage while the longer ones allow for more creativity. Tyrkkö neatly sets out his statistical reasoning and applicability of particular tests to the proposed research questions – the discussion can well serve as a short tutorial in statistics for historical linguistics itself.

María José López-Couso and Belén Méndez-Naya give a diachronic overview of the use of *think*- and *seem*-parentheticals in order to question the proposal that these constructions originated from matrix clauses in complement structures. Their data support an alternative scenario, whereby adverbial clauses *so/as* ... *think/seem* give rise to parentheticals. The authors present a useful overview of the types of constructions with *think* and *seem* in the Helsinki Corpus, and a data-driven exploratory categorization. Perhaps the discussion could be enriched with statistical significance tests regarding the shifts in the preference for complement and parenthetical constructions over time.

In the next chapter, Timothy Coleman and Dirk Noël take a comparative perspective on the diachronic development of the constructions (*consider*) *sb* to be (ACI) and *be supposed to* (NCI) in English and Dutch, which offers a counterexample

to the path of grammaticalization from deontic to epistemic meanings. They address the theory that both ACI and NCI constructions were modelled on Latin and review earlier ideas on the development of *be supposed to* in English. In fact, the discussion of the English constructions is more solid in the paper. The Dutch corpus sources are quite heterogeneous: different genres from different periods, with a temporal span from 1640 to 1998 and large discrepancies in word count. This mish-mash of material may, in a way, provide illustrative context for the employment of the constructions in question but it does not allow for meaningful comparisons across genres or periods, as the word counts are not normalized. The findings, modest in counts as they are, suggest that the deontic meaning appeared in Dutch only in the twentieth century, which – as the authors suggest – was due to “unprecedented potential for contact influence . . . because of the enormous increase in the ‘consumption’ of English by speakers of Dutch” (p. 229). Although the contact theory is plausible, there is too little contextualization for this statement to accept it without reservation.

An interesting take on corpus normalization can be found in Jan K. Lindström’s chapter on fronted negation in spoken Swedish. Since this is a syntactic phenomenon characteristic of spoken registers, the author has chosen to compare counts across periods relative to the number of sentences in a corpus of drama dialogue (1725–2000). The findings support the theory that peripheral dialects preserve more archaic features – in this case Finland Swedish retains features of speech that are further strengthened by language contact, while standard Swedish develops differently. The qualitative discussion of interactional factors behind the fronted negation is quite persuasive but the suggestion that older texts score lower because they “do not reflect natural speech very well” (p. 244) should be explored further as it may undermine the study design and the choice of drama as a reflection of spoken discourse.

The third part of the book, on utterances and dialogues, opens with a chapter on compliments by Andreas H. Jucker and Irma Taavitsainen. It is a sequel to their earlier work (2008), where British data (1500–1800) were surveyed; now the authors turn their attention to American English (1810–2010), taking their approach further in a quantitative direction. Anyone interested in speech acts will benefit from a thorough overview of this frequently studied pragmatic phenomenon given here. With focus on personal and ceremonious compliments, the authors critically assess various approaches to collecting and retrieving speech acts: from introspective research to corpus studies (IFIDs, syntactic frames, metacommunicative expressions). For instance, it seems that earlier diary-based research on compliments was skewed because of the gender of the data collector. This study serves to showcase sound methodological decisions: random sampling to limit the data for qualitative assessment, coding the data for several well-defined factors and coder agreement assessment prior to data inclusion. One of the most striking findings is that for the last 200 years American men have given and received compliments more frequently than women. The next step could be to correlate these results with the topic of the compliment in a less speculative way. The authors were also able to confirm that American culture is geared more towards accepting compliments, which has been a growing strategy at least since the 1820s.

The study on verbal aggression by Dawn Archer is a testing ground for a new automatic semantic annotation system (USAS). This tool is used to identify the “aggression space” in late eighteenth-century trial records. Bearing in mind various methodological pitfalls, the author carefully introduces a hierarchically organized system of semtags and explains how they can be used to construct “pragmatic space” (Jucker and Taavitsainen, 2000) in historical texts. In her discussion of verbal aggression, the author employs a useful face aggravation scale (Archer, 2011), which captures intentional and incidental FTAs, quite characteristic of courtrooms. I find the discussion of the automatic corpus annotation and extraction very detailed, honest and therefore illuminating: it is indeed problematic that out of all semtags the tagger captured only ‘speech acts’ and ‘unknown’ categories in a more pronounced way, and that the numbers of aggression-related semtags were low in court proceedings. The author identifies issues for further investigation, e.g. aggression being veiled or emerging from a longer passage, and for refinement, e.g. spelling variation and its influence on automatic tagging.

Finally, Angela Schrott approaches Old Spanish counselling dialogues (c. 900–1400) from a function-to-form perspective. She redesigns Eugenio Coseriu’s model of “language as a cultural competence” (p. 304) for a historical pragmatic application, trying to create an interface between levels of language use and related types of knowledge. I found this attempt interesting but lacking in precision. It is not surprising that a medieval knight on a battlefield and a protagonist of a courtly romance written 150 years later would seek counsel in very different ways – was the author’s sole aim an illustration of these different counselling strategies? In the quantitative part of the chapter, Schrott uses Treecloud to capture collocation clouds emerging in counselling contexts in both texts. The results of a log-likelihood test applied to lexeme frequency in counselling and non-counselling contexts reveal the importance of metapragmatic expressions, which speaks in favour of the methodology employed by Jucker and Taavitsainen in their chapter.

The editors have ensured that each paper be written in a similar structural frame, which facilitates comparing the premises and results across the studies in the book. They also made sure that the focus was clearly diachronic. All papers are carefully signposted and the combination of qualitative and quantitative methods permeates the whole volume. It becomes clear that it is a real challenge to do both in an equally expert manner. Several chapters meet high standards on both grounds (Claridge and Kytö, Tyrkkö, Jucker and Taavitsainen). The remaining chapters are much better on the qualitative side, bringing to the fore the context for a given pragmatic aspect of linguistic behaviour and the function it plays. The current turn towards ‘big’ and ‘dirty data’ poses significant problems. In some studies, e.g. Norlund and

Pekkarinen's or Küngas's, the approach to data is meticulously documented but some methodological decisions seem to fall short of the 'big data' challenges. Much more convincing are contributions such as Ricca and Visconti's or Archer's where the authors are aware of the shortcomings of a large imbalanced corpus or 'dirty data' and at least speculate how these problems could be addressed in future research. Another issue with which some authors (e.g. Shinzato) have dealt better than others is discussing spoken language features on the basis of written materials. The lack of engagement with any particular analytic framework which justifies the link between speech and writing is especially striking in Lindström's chapter.

In general, the volume has many strengths and will appeal to pragmaticists with various interests: from an impressive range of languages under analysis (it is a signal for Slavonic linguistics to make its contribution), through intriguing historical and contextual detail, testing new corpus tools and statistical approaches, to redefining and delineating the scope of diachronic pragmatics itself. Reading through the chapters, I noticed a lot of potential for cross-linguistic projects and collaborations – other readers will surely feel similarly inspired by this timely, well-organized and original research collection.

References

- Archer, Dawn, 2011. [Cross-examining lawyers, facework and the adversarial courtroom](#). *J. Pragmat.* 43 (13), 3216–3230.
- Biber, Douglas, et al. (Eds.), 1999. [Longman Grammar of Spoken and Written English](#). Longman, London.
- Claridge, Claudia, Arnovick, Leslie, 2010. [Pragmaticalisation and discursisation](#). In: Andreas, H. Jucker, Irma, Taavitsainen (Eds.), *Historical Pragmatics*. De Gruyter Mouton, Berlin, pp. 165–192.
- Jucker, Andreas H., Taavitsainen, Irma, 2000. [Diachronic speech act analysis: Insults from flyting to flaming](#). *J. Hist. Pragmat.* 1 (1), 67–95.
- Taavitsainen, Irma, Jucker, Andreas H., 2008. ['Methinks you seem more beautiful than ever': Compliments and gender in the history of English](#). In: Andreas, H. Jucker, Irma, Taavitsainen (Eds.), *Speech Acts in the History of English*. John Benjamins, Amsterdam/Philadelphia, pp. 195–228.

Joanna Kopaczyk is a post-doctoral research assistant at the University of Edinburgh. Her recent books include a study of standardising patterns in early legal discourse (*The Legal Language of Scottish Burghs. Standardisation and lexical bundles (1380–1560)*, Oxford University Press, 2013), and the first volume to explore the applications of the notion of a 'community of practice' to historical linguistic research (*Communities of Practice in the History of English*, John Benjamins, 2013, co-edited with Andreas H. Jucker). She also holds an assistant professorship at Adam Mickiewicz University in Poznań.

Joanna Kopaczyk^{a,b,*}

^aAngus McIntosh Centre for Historical Linguistics, The University of Edinburgh, EH8 9AD UK

^bFaculty of English, Adam Mickiewicz University in Poznań, Poland

*Correspondence to: Angus McIntosh Centre for Historical Linguistics, PPLS, DSB 2.01,
The University of Edinburgh, EH8 9AD UK. Tel.: +44 (0) 131 651 5507.

E-mail address: joanna.kopaczyk@ed.ac.uk